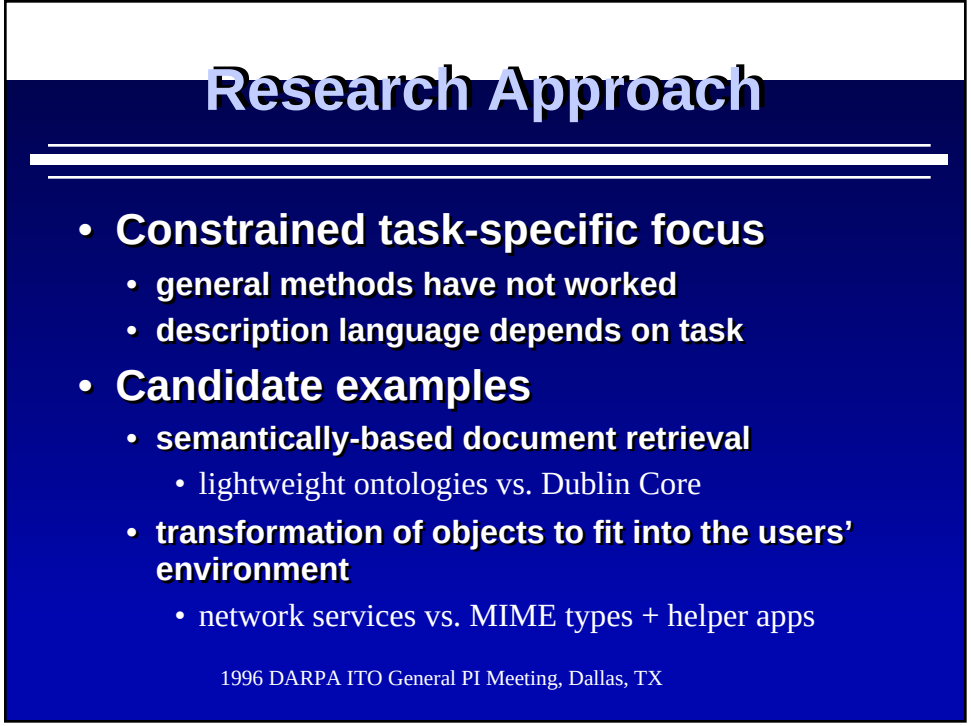
Our working group identified four principal opportunities for research in the area of meta-data for self-describing objects. The slide lists these. Our emphases were determined by considering a wide variety of uses for meta-data and self-description. From these scenarios of use, we determined a primary need for users or programs to be able to describe objects and have these descriptions support a variety of tasks. These tasks ranged from search to transformation to interactive presentation, so that objects might be found and employed to address a variety of user requirements.

Given limited time in our workshop, we were able to consider to some depth the first two of these four principal topics, namely “Languages and methods for describing objects” and “Generation of descriptions.” The other two focus areas, “Intelligent object search and apply” and “Languages for expressing user needs,” received less attention.

The remainder of this talk focuses on the first two topics, which are highlighted in yellow on the slide.



Much consideration of abstract topics such as “meta-data” seems to founder when it is ungrounded in some particular task or usage scenario. Our working group felt that an opportunity was apparent to pursue the focus topics if the work could be focused in some particular tasks. The argument for this was two-fold. First, general methods of description, semantics, and object-oriented systems have not produced useful results. The general concepts and methods seem too weak to contribute significant progress or useful results. Second, any pursuit of a language for describing objects would appear to depend heavily upon a notion of task and purpose. Languages without some notion of pragmatics were not deemed of much utility.

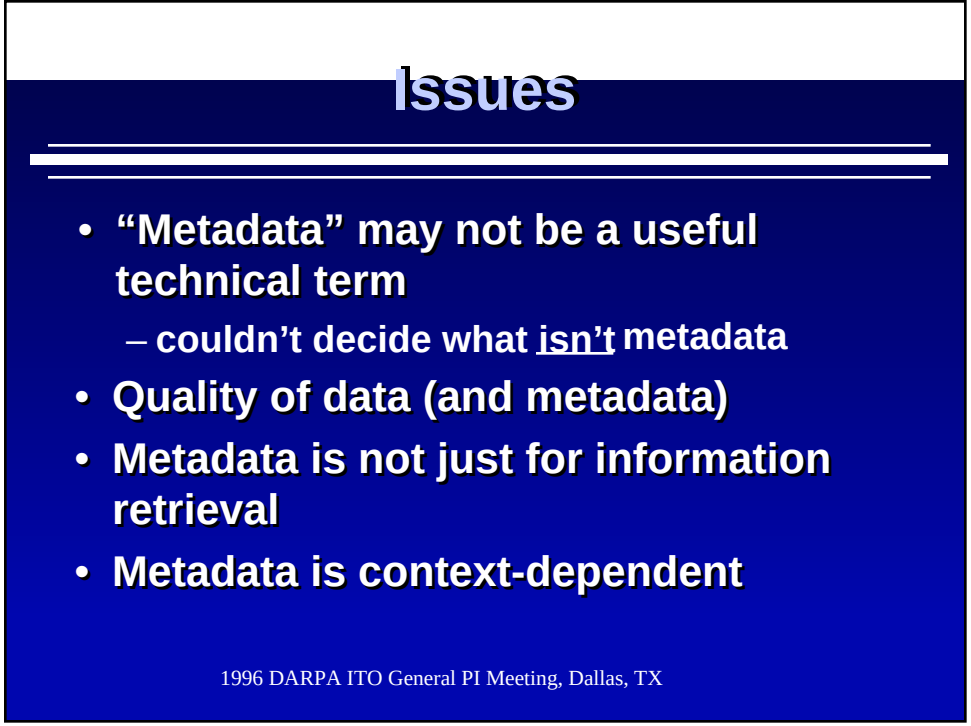
We identified two examples of tasks that could provide useful constraints for this research area. First, the task of retrieving documents based on semantic content was judged attractive. In contrast with library science approaches, such as the “Dublin Core,” the working group members felt that significant progress could be made with simplified or “lightweight” ontologies. For example, both the attributes of “Author” and “Date” were considered ripe for improved semantic processing, because each can be characterized readily with relations that could exploit rule-based processing for matching and queries.

The second example task was one of transforming objects to fit the constraints imposed on the user's particular computing environment. Wheras current systems have a notion of MIME types that point to users’ helper applications, we felt that access to a wide variety of objects and their methods could be

The group identified a number of promising approaches, five of which are listed on this slide. We believe that considerable improvement in functionality can be provided in various tasks through the use of simplified or “lightweight” ontologies augmented with semantic and pragmatic knowledge employed to automate key tasks. This relates to the second idea, that of executable or interpretable metalanguages. In this case, we believe that it would be an improvement to adopt some formalisms for describing objects and their behaviors that would allow these behaviors to be executed in whatever context the object was accessed. These meta-languages would go beyond simple attribute-value descriptions, providing capability for the objects to render, transform, or otherwise engage in interactions with clients.

Because characterization and description are important and time-consuming tasks, it seemed promising to address these tasks themselves through potential new combinations of methods that have previously been used primarily in isolation. Specifically, we thought that statistical and knowledge-based characterization approaches could be combined, and new efforts at data mining might be brought to bear on the problem as well. Furthermore, because some of the best descriptions are created by humans, we thought it a promising avenue to explore ways to reward people for creating successful descriptions, perhaps as a by-product of their other uses of objects.

Finally, descriptions are intended to make subsequent exploitations of objects easier, faster, and effective. Thus, it would seem valuable to establish a



The working group grappled with several implicit issues underlying the topic and different orientations towards it. Among the many assumptions and issues, four particular ones stood as important. First, the technical terminology for this area of research is immature, and the term "metadata" meant many different things. It seemed like metadata in one view was data in another, and data in one view might be metadata in another.

The quality of data was deemed an important metadata issue, but one which seems broad, deep and difficult. The quality of data is a reflection of limits in collection and processing methods, errors in communication and summarization, and pragmatic effectiveness, among other issues. No simple way of treating "quality" was apparent.

There are many uses of metadata. We wanted to be sure that people did not think metadata is exclusively useful for retrieval. Moreover, we discussed a range of kinds of objects, from information, to programs, to dynamic processes, and metadata was potentially useful in describing, accessing, or exploiting any of these.

Some participants were primarily interested in the use of metadata for controlling access and monitoring use of "intellectual property" contained inside the corresponding object.

Any such particular focus requires that the metadata characterize the aspects of the objects of principal import to the intended purpose. In short, the "right"